

# On The Non-Gaussian Errors in High-z Supernovae Type Ia Data

Meghendra Singh<sup>1</sup>, Ashwini Pandey<sup>2</sup>, Amit Sharma<sup>2</sup>, Shashikant Gupta<sup>2</sup>, Satendra Sharma<sup>3</sup>

<sup>1</sup> Dr.A.P.J.Abdul Kalam Technical University, Uttar Pradesh, Lucknow 226021, India;  
*meghendrasingh\_db@yahoo.co.in*

<sup>2</sup> Amity University Haryana, Gurgaon, Haryana 122413, India.

<sup>3</sup> Yobe State University, Damaturu, Yobe State, Nigeria.

Received— ; accepted —

**Abstract** The nature of random errors in any data set is Gaussian is a well established fact according to the Central Limit Theorem. Supernovae type Ia data have played a crucial role in major discoveries in cosmology. Unlike in laboratory experiments, astronomical measurements can not be performed in controlled situations. Thus, errors in astronomical data can be more severe in terms of systematics and non-Gaussianity compared to those of laboratory experiments. In this paper, we use the Kolmogorov-Smirnov statistic to test non-Gaussianity in high-z supernovae data. We apply this statistic to four data sets, i.e., Gold data(2004), Gold data(2007), Union2 catalogue and the Union2.1 data set for our analysis. Our results shows that in all four data sets the errors are consistent with the Gaussian distribution.

**Key words:** cosmology: Data Analysis, Statistics and Probability

## 1 INTRODUCTION

The light curves of Type Ia supernova (SNIa) have been used as cosmological distance indicators ([Riess et al. 1998](#); [Perlmutter et al. 1999](#)) to mark out the expansion history and to detect cosmic acceleration as well. The overall picture of the Universe is consistent with a model known as the  $\Lambda$ CDM, consisting of around one quarter of baryonic and dark matter and three quarters of dark energy. The dark energy can be treated as a cosmic-fluid with equation of state  $P = w\rho$ ; where the pressure ( $P$ ) is allowed to be negative. The SNIa data can be used to constrain the equation of state parameter ( $w$ ) which is the key to study dark energy ([Freedman et al. 2009](#); [Hicken et al. 2009](#); [Rest et al. 2014](#); [Scolnic et al. 2014](#)) .

However, many alternative explanations exist for dark energy and its exact nature is also unknown. For instance, a classical fixed cosmological constant,  $\Lambda$ , yields  $w = -1$ , whereas other models (e.g. quintessence) yield values of  $w > -1$  ([Huntere et al. 2001](#)). To overcome this difficulty, precise enough data is required to detect fluctuations in the dark energy. The data should also cover wide range of redshifts

is obtained by the observations of the SNIa. Determination of supernovae distances having high precision and tiny systematic errors is crucial for above purpose; and we would like to be certain that their statistics is well understood. Further, if Central Limit Theorem holds (Kendall et al. 1977), the statistical uncertainties in SNIa data should follow Normal distribution. The systematics, if present, have to be identified and removed separately. Treatment of the errors becomes more important in astronomy since it is hard to repeat or perform the experiments in controlled way unlike the laboratory experiments. In the present paper, we use the Kolmogorov-Smirnov test (hereafter KS test) in an elegant way to detect the non-Gaussian uncertainties in SNIa data.

This paper aims to address the above mentioned problems. The rest of the paper is formed as follows: In Section 2, we illustrate the different data sets used for our analysis, while Section 3 contains detailed description of methodology used. In Section 4, we continue and put forward our results for various data sets and lastly Section 5 is reserved for conclusions.

## 2 DATA

The Gold data GD04 (Riess et al. 2004) containing 157 SNe, GD07 (Riess et al. 2007) containing 182 SNe along with the more recent and larger data sets Union2 (Amanullah et al. 2010) and Union2.1 (Suzuki 2012) containing 557 and 580 SNe respectively are used to carry out our investigation. The redshift  $z$  and the distance modulus  $\mu$  are the measured quantities in the data. If  $m$  is the apparent magnitude and  $M$  is the absolute magnitude, then distance modulus is defined as:

$$\mu(z) = m(z) - M, \quad (1)$$

The apparent magnitude  $m(z)$  and hence distance modulus  $\mu(z)$  depends on the intrinsic luminosity of a supernova, its redshift  $z$  and the cosmological parameters. The distance modulus  $\mu(z)$  and the luminosity distance  $d_L$  are related as:

$$\mu(z) = 5 \log(d_L(z)) + 25, \quad (2)$$

where the luminosity distance is measured in  $Mpc$  and follows:

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{dx}{h(x)}, \quad (3)$$

where  $h(z; \Omega_M, \Omega_X) = H(z; \Omega_M, \Omega_X)/H_0$ , and hence it is independent of  $H_0$  but depends purely on densities of dark matter ( $\Omega_M$ ) and dark energy ( $\Omega_X$ ). The variation of  $\Omega_X$  with redshift is already encoded in the cosmological models; for instance  $\Omega_X$  is a constant in the  $\Lambda$ CDM model. The nature of relation of  $\mu$  with  $M$  is linear, however, that with luminosity distance is logarithmic. This implies the logarithmic dependence of  $\mu$  on Hubble parameter  $H_0$  as well.

## 3 METHODOLOGY

We now give an introduction to the method of our analysis. Originally it was presented in Singh et al. 2015(hereafter GS15) to find non-Gaussianity in the HST Key Project Data.

If the correct theoretical value of the distance modulus of  $i^{th}$  supernova at redshift  $z$  is  $\mu_i^{th}(z)$ , then the observed value  $\mu_i^{obs}$  will be

where  $\sigma_i$  is the error in the measurement of distance modulus. We expect these errors to be completely random, however, there could be some undesired contribution from systematic effects. For the time being we assume that the systematic part in the errors is zero. We show in next paragraph that the presence of systematic errors will not affect our analysis. Further, Central Limit Theorem suggests that the random part of the errors should be Gaussian in nature with mean value zero. Now we define a quantity  $\chi_i$  such that:

$$\chi_i = \frac{\mu_i^{obs} - \mu_i^{th}(z)}{\sigma_i}, \quad (5)$$

Clearly  $\chi_i$  should follow the standard normal distribution  $N(0,1)$ , i.e., Gaussian distribution with zero mean and unit variance. The effect of random errors is to scatter the data around the true value and that of systematics is to shift the average away from the true value. If the systematics are present they will just shift the average, hence one should subtract the best-fit value rather than true theoretical value in Eq. 5. Thus Eq. 5 takes the following form for a given SN:

$$\chi_i = \frac{\mu_i^{obs} - \mu_i^{best\,fit}(z)}{\sigma_i}, \quad (6)$$

where  $\mu_i^{best\,fit}(z)$  is calculated using the best-fit values of cosmological parameters. Statistical independence among supernovae in our analysis is an obvious assumption.  $\chi_i$  defined in Eq. 6 should follow a standard normal distribution, i.e., Gaussian with zero mean and unit standard deviation.

We use the flat  $\Lambda$ CDM cosmology in our analysis, since it fits the SNe data well. However, other cosmological models could also be investigated using a similar approach. In order to get best-fit values of Cosmological parameters we minimize  $\chi^2$ , which is defined as:

$$\chi^2 = \sum_{i=1}^N \left[ \frac{\mu_i^i - \mu^{\Lambda CDM}}{\sigma_i} \right]^2, \quad (7)$$

Once again we emphasize that, Eq. 7 is used to find the best-fit values of cosmological parameters and it is then used in Eq. 6 to calculate  $\chi_i$ .

As argued earlier,  $\chi_i$  should follow standard normal distribution. To check this, we use KS test to determine whether or not a given sample follows the Gaussian distribution (Press 2007). For this we set our null hypothesis as: "The errors in the SNe data are drawn from a Gaussian distribution". Thus  $\chi_i$ 's in Eq. 6 would follow standard normal distribution. We apply KS test to calculate the test statistic and the p-value which is the probability of attaining the observed sample results when the null hypothesis is true.

For this, we use Matlab function  $kstest[h,p,k,cv]$ ; where:  $p$  represents the probability of the data errors being drawn from Gaussian distribution,  $k$  is the maximum distance between the two distributions (CDF), and  $cv$  is the critical value which is decided by the significance level ( $\alpha$ ). Different values of  $\alpha$ , indicate different tolerance levels for false rejection of the null hypothesis. For instance,  $\alpha = 0.01$  means that we allow 1% of the times to reject the null hypothesis when it is true.  $cv$  is the critical value of the probability to obtain/generate the data set in question given the null hypothesis; and is to be compared with  $p$ . A value  $h = 1$  is returned by the test if  $p < cv$  and the null hypothesis is rejected. While for  $p > cv$ ,  $h$  remains 0

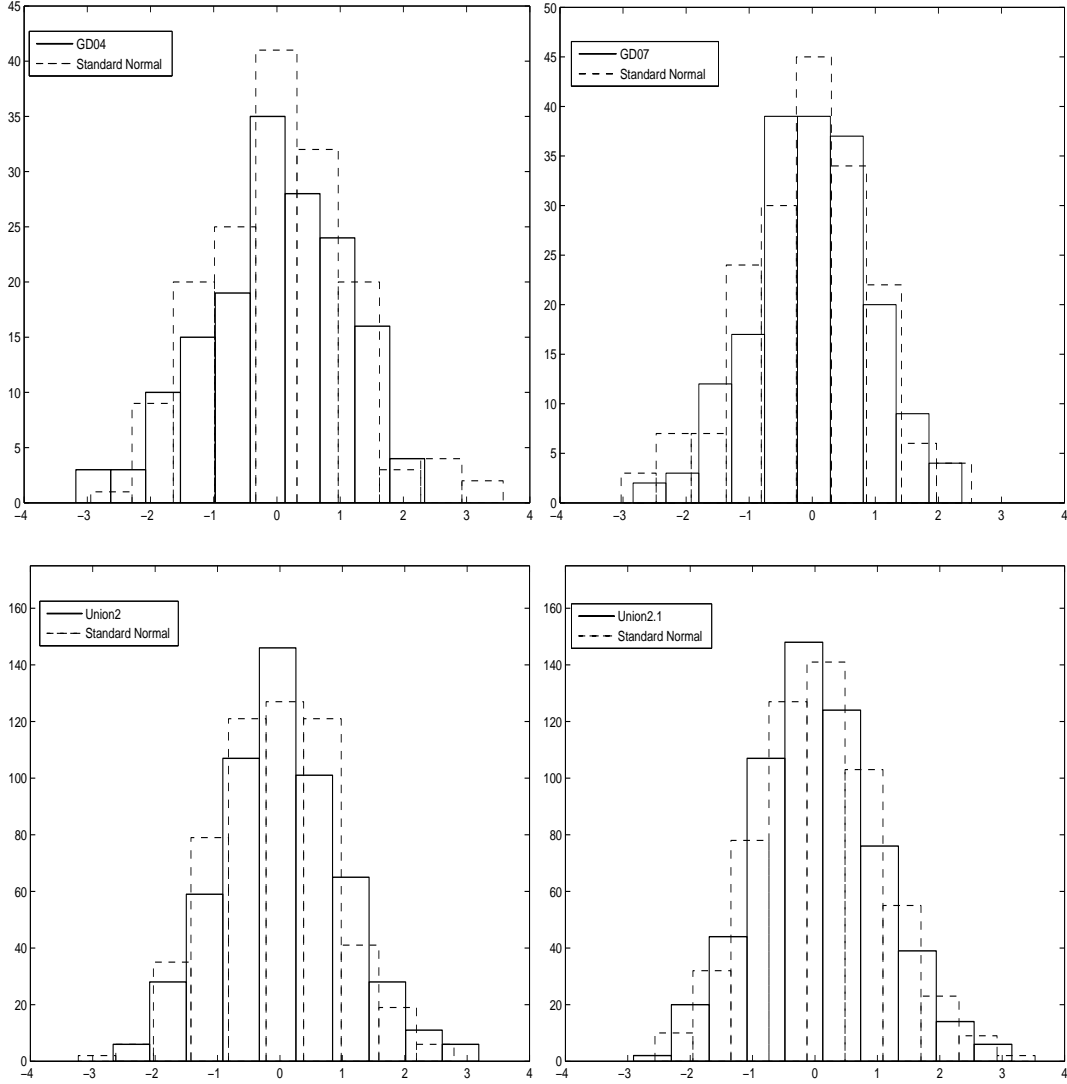


Fig. 1: Histogram of  $\chi_i$  for various data set is compared with that of standard normal distribution.

Table 1: The best-fit values for various data sets.

Data Set	# SNe	$\Omega_M$	$H_0$	$\chi^2/dof$
GD04	157	0.30	64.5	1.143
GD07	182	0.33	63.0	0.883
Union2	557	0.27	70.0	0.975
Union2.1	580	0.28	70.0	0.973

## 4 RESULTS

We apply the statistic discussed in section 3 on various SNe data sets and present the results here. Similar analysis was presented in [Gupta et al. 2010](#) (hereafter GS10) and in [Gupta et al. 2014](#) (hereafter GS14) using a different method ( $\Delta\chi^2$ ) based on extreme value theory.

As a first check, we calculate the best-fit values of cosmological parameters for all four data sets by

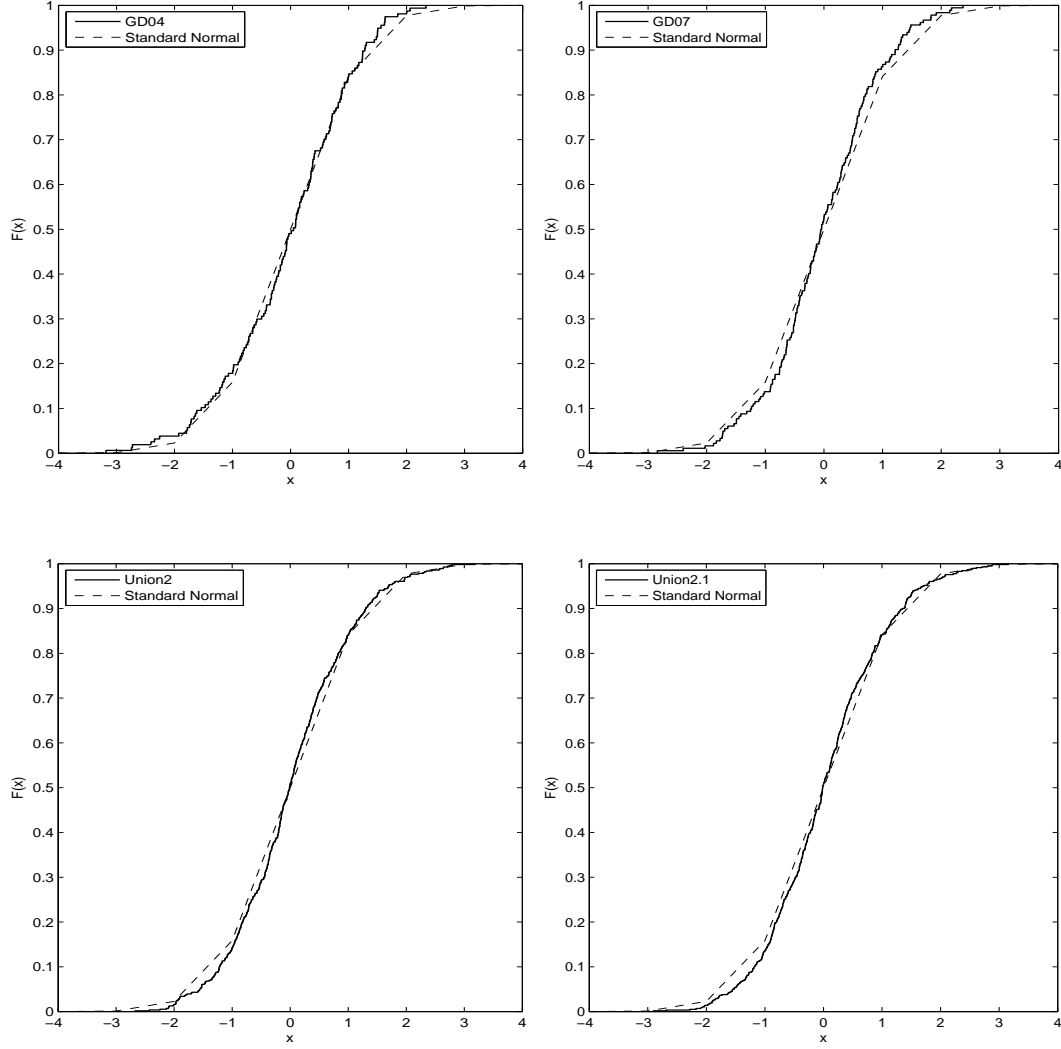


Fig. 2: Comparison of cumulative distribution of  $\chi_i$  for different data sets with their Gaussian CDF. Smooth curve represents the Gaussian CDF

( $\Omega_M$ ) and consequently smaller expansion rate ( $H_0$ ) compared to Union2 and Union2.1. One important fact is that the  $\chi^2$  per degree of freedom secures smallest value for GD07 while largest for GD04, indicating the overestimation and underestimation of errors in GD07 and GD04 respectively.

We calculate  $\chi_i$ 's as defined in Eq 6 for each data set using the best-fit values presented in Table 1. Further, we generate four sets of random numbers following Gaussian distribution with zero mean and unit standard deviation. Fig. 1 represents the comparison of histograms of Gaussian random numbers with that of  $\chi_i$ 's of each data set.

Secondly, the result of the KS test which is arrived at by comparing the calculated cumulative distributions for  $\chi_i$ 's with that of Gaussian distribution are presented in Table 2. The second, third and fourth column in Table 2 denote values of  $p$ ,  $k$  and  $cv$  respectively. Since  $p > cv$  in all cases giving  $h = 0$ , which means that we can not reject the null hypothesis that the errors are drawn from a Gaussian distribution; is

Table 2: Results of KS test for various data sets.

Data Set	$p$ value	$k$	$Cv$
GD04	0.9280	0.0425	0.1073
GD07	0.7872	0.0475	0.0997
Union2	0.7328	0.0288	0.0572
Union2.1	0.6764	0.0296	0.0561

## 5 CONCLUSIONS

We have used the method presented in GS15 to detect non-Gaussianity in the error bars in Supernovae data. Our main conclusions for this part of our work are following: (a) The errors are probably underestimated in GD04 and overestimated in GD07. In this sense, both the sets stand on extreme positions. (b) For a flat  $\Lambda$ CDM cosmology GD07 favors slightly higher matter density and this can be verified by the fact that in GD07 the distances are smaller compared to that in GD04 set for common supernovae. (c) Comparison with GS10 and GS14: GD04 was shown to have non-Gaussian component of errors while in KS test, it shows the highest probability of being consistent with Gaussian distribution. (d) The hypothesis that the errors are drawn from Gaussian distribution can not be rejected for all the data sets discussed in the present paper.

**Acknowledgements** Meghendra Singh thanks DMRC for Support, Shashikant Gupta thanks Tarun Deep Saini for discourse; and compeers of Amity School of Applied Science for eternal assistance. Authors thank to the anonymous reviewer for valuable suggestions.

## References

- Amanullah, R., et al. 2010, ApJ, 716, 712.
- Freedman, W. L., et al. 2009, ApJ, 704, 1036.
- Gupta, Shashikant, & Saini, Tarun Deep., 2010, MNRAS, 407, 651.
- Gupta, Shashikant, & Singh, Meghendra., 2014, MNRAS, 440, 3257.
- Hicken, M., et al. 2009, ApJ, 700, 1097.
- Huntere,D., et al. 2001, Phys. Rev. D, 64, 123527.
- Kendall, M., & Stuart, A., 1977, London: Griffin,4th ed.
- Perlmutter, et al. 1999, ApJ, 517, 565.
- Press, W. H., Numerical recipes, The art of scientific computing, Cambridge University Press, 2007.
- Rest, A., et al. 2014, ApJ, 795, 44.
- Riess, A. G., et al. 1998, AJ, 116, 1009.
- Riess, A. G., et al. 2004, ApJ, 607, 665.
- Riess, A. G., et al. 2007, ApJ, 659, 98.
- Scolnic, D., et al. 2014, ApJ, 795, 45.
- Singh, Meghendra, Gupta, Shashikant, Pandey, Ashwini, Sharma, Satendra, 2015, arXiv:1506.06212.
- Suzuki, N., et al. 2012, ApJ, 746,85.